

Varying cluster sizes in trials with clusters on one treatment arm: sample size adjustments when testing treatment effects with linear mixed models.

Citation for published version (APA):

Candel, M. J. J. M., & van Breukelen, G. J. P. (2009). Varying cluster sizes in trials with clusters on one treatment arm: sample size adjustments when testing treatment effects with linear mixed models. *Statistics in Medicine*, 28, 2307-2324. <https://doi.org/10.1002/sim.3620>

Document status and date:

Published: 01/01/2009

DOI:

[10.1002/sim.3620](https://doi.org/10.1002/sim.3620)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 05 May. 2023

Varying cluster sizes in trials with clusters in one treatment arm: Sample size adjustments when testing treatment effects with linear mixed models

Math J. J. M. Candel^{*,†} and Gerard J. P. Van Breukelen

*Department of Methodology and Statistics, Maastricht University, P.O. Box 616, 6200 MD Maastricht,
The Netherlands*

SUMMARY

Trials in which treatments induce clustering of observations in one of two treatment arms, such as when comparing group therapy with pharmacological treatment or with a waiting-list group, are examined with respect to the efficiency loss caused by varying cluster sizes. When observations are (approximately) normally distributed, treatment effects can be estimated and tested through linear mixed model analysis. For maximum likelihood estimation, the asymptotic relative efficiency of unequal versus equal cluster sizes is derived. In an extensive Monte Carlo simulation for small sample sizes, the asymptotic relative efficiency turns out to be accurate for the treatment effect, but less accurate for the random intercept variance. For the treatment effect, the efficiency loss due to varying cluster sizes rarely exceeds 10 per cent, which can be regained by recruiting 11 per cent more clusters for one arm and 11 per cent more persons for the other. For the intercept variance the loss can be 16 per cent, which requires recruiting 19 per cent more clusters for one arm, with no additional recruitment of subjects for the other arm. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: asymptotic relative efficiency; clustering effects of treatments; individually randomized group treatment; sample size; varying cluster sizes

1. INTRODUCTION

Many study designs evaluating the effect of an intervention are characterized by observations being correlated within clusters. This may arise in group or cluster randomized trials [1, 2], where groups are assigned to one of several treatment conditions. Examples are cities with all of its citizens being assigned to a prevention program or not, or patients nested within general practices being assigned to one of several medication treatments. Groups in these cases are the units of assignment.

*Correspondence to: Math J. J. M. Candel, Department of Methodology and Statistics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

†E-mail: Math.Candel@Stat.unimaas.nl

Clustering of observations may also arise when individuals instead of groups are the units of assignment. This is the case when the treatment itself induces clustering such as when treatments are given to groups of individuals [3–5]. These designs are known as individually randomized group treatment trials [6]. In such designs, interactions between persons within a group may lead to the observations on the outcome variable being correlated. The clustering may occur in only one of the treatment arms, such as when group therapy is compared with a condition involving only medication (e.g. [7–9]) or when group therapy is compared with a condition not involving any kind of intervention (e.g. [5, 10–12]).

Clustering effects due to treatment may also occur if the treatment is not given to a group as a whole but on an individual basis. If several patients are treated by the same therapist, then patients will be treated in a more similar way than patients treated by different therapists. Such therapist effects may also lead to correlated observations within clusters [6, 13]. In addition, in this case clustering may occur in only one of two treatment arms, such as when a comparison is made with a waiting-list condition (e.g. [14–16]) or with a pharmacological or a placebo condition (e.g. [17]).

In the present study we will consider designs with treatment-induced clustering in one of two treatment arms. Formulas for calculating the sample size for such designs have been derived by Moerbeek and Wong [18] and Roberts and Roberts [4]. The present study extends these studies by examining unequal cluster sizes. Unequal cluster sizes may be due to variation in actual cluster size, but also due to non-response or dropout of subjects and therefore is quite common. We will examine the efficiency loss for estimating the treatment effect due to variation of cluster size. We assume that the data within each condition are (approximately) normally distributed and are analyzed with linear mixed models. The efficiency loss will be derived for the asymptotic case, when estimating the treatment effect through maximum likelihood (ML), and it will be checked for finite samples by an extensive Monte Carlo simulation study. The issue is how much power is lost and how to compensate for this loss.

The number of clusters as encountered in relevant studies as well as the size of the clusters themselves are small (e.g. [4, 6, 8, 12, 19, 20]). The Monte Carlo study will therefore examine the accuracy of the asymptotic efficiency loss for small sample sizes. The accuracy will be examined for both ML and restricted ML estimation.

Although the primary interest is usually in the treatment effect, the variance of the random effects in the linear mixed model may also be of interest. The treatment may be more successful for some clusters than for others. Variations between clusters in the treatment effect may be due to differences in the interactions between the clusters' members or due to the influence of therapists assigned to clusters. For the model under consideration, the random effects reflect such variations between clusters in the effectiveness of the treatment. The variance of the random effects can be used to construct prediction intervals for these cluster-specific treatment effects [21, 22]. We will therefore also examine the efficiency loss due to varying cluster sizes when estimating the variance of the random effects.

More complicated clustering is possible, such as when a condition with group therapy is compared with a condition without any intervention, and instead of individuals, mental health centres are assigned to the treatment arms. In these cases there may be clustering at two levels within one arm (persons being nested within therapy groups, which in turn are nested within mental health centres), and clustering at one level within the other arm (persons being nested within mental health centres). Then compared with the design that we consider in this paper, there is one additional level of clustering in each arm. Such clustering will not be considered in the present study.

The structure of the paper is as follows. Section 2 presents the linear mixed effects model for trials comparing a condition with clustering to a condition without clustering. Section 3 provides explicit expressions for the asymptotic relative efficiency (RE) of the estimator of the treatment effect and of the random intercept variance in such designs. Section 4 discusses the design and results of a Monte Carlo simulation that examines the RE of unequal versus equal cluster sizes for various cluster size distributions with realistic cluster sizes. Section 5 illustrates how the results of the present study can be used in planning a trial. The paper closes with some implications for the planning phase of trials as well as some issues for further research.

2. SPECIFICATION OF THE LINEAR MIXED EFFECTS MODEL

In the condition with clustering, called the cluster condition in what follows, we have K clusters (e.g. K therapy groups). In cluster j ($j = 1, \dots, K$) there are n_j persons and the total number of persons in the cluster condition is $N = \sum_{j=1}^K n_j$. In the other treatment condition, henceforth called the control condition, there are no clusters of correlated observations, and the number of persons in this $K+1$ th cluster is n_{K+1} . In the case of equal cluster sizes, we have $n_j = n$ for $j = 1, \dots, K$, but in general not for $j = K+1$ (e.g. when $n_{K+1} = N$). The dependent variable is a quantitative outcome, denoted as y_{ij} for person i in cluster j ($j = 1, \dots, K+1$). If in each condition y_{ij} is (approximately) normally distributed, the linear mixed effects model is an adequate tool for data analysis. The corresponding analysis model is then as follows (cf. [4, 5, 18]):

$$y_{ij} = \beta_0 + (\beta_1 + u_{0j} + \varepsilon_{ij})Int_{ij} + \delta_{ij}(1 - Int_{ij}) \quad (1)$$

where Int_{ij} denotes the treatment condition for person i in cluster j , and is coded as 1 for persons in the cluster condition and 0 for persons in the control condition. With this coding scheme, β_0 represents the mean score of the condition without clusters and β_1 represents the treatment effect (such as the treatment effect of group therapy versus pharmacological treatment). The terms ε_{ij} and u_{0j} represent a random person and random cluster effect in the cluster condition, which are assumed to be independently normally distributed with variances σ_ε^2 and σ_0^2 , respectively. The random effect for persons in the control condition, δ_{ij} , is also independently normally distributed with a possibly different variance σ_δ^2 . Thus, the model has five unknown parameters that have to be estimated: two fixed regression weights, β_0 and β_1 , and three variance components, σ_0^2 , σ_ε^2 and σ_δ^2 . The model parameters can be estimated through ML [22].

Commonly the parameter β_1 , representing the treatment effect averaged across all clusters, is the parameter of primary interest. In case persons are randomly assigned to clusters (e.g. as in group therapy), we can consider $\beta_1 + u_{0j}$ as the treatment effect for cluster j ($j = 1, \dots, K$). Since u_{0j} is normally distributed, 95 per cent of these cluster-specific treatment effects will be between the boundaries $\beta_1 - 1.96\sigma_0$ and $\beta_1 + 1.96\sigma_0$ [21, 22]. This interval has been denoted as the predictive interval for the cluster effects (not to be confused with the confidence interval for the treatment effect β_1) [21]. Hence, for expressing the variation in effectiveness of the treatment across clusters by a predictive interval, the random intercept variance σ_0^2 is also a parameter of interest.

A relevant concept is the intraclass correlation. This is the correlation between the outcome measures for two randomly drawn persons from the same cluster, and thus measures the dependency among observations taken on members within the same cluster. For the treatment arm with clustering, the intraclass correlation is: $\rho = \sigma_0^2 / (\sigma_0^2 + \sigma_\varepsilon^2)$. The larger the cluster variance as

compared with the person variance in the cluster condition, the larger the intraclass correlation. The correlation for two randomly drawn persons from two different clusters in this treatment arm is zero. This treatment arm thus consists of clusters of correlated observations. For any two randomly chosen persons in the other treatment arm, there is no correlation between outcome measures.

3. ASYMPTOTIC RE

We are interested in the estimator of the treatment effect, β_1 , and will also consider the estimator of σ_0^2 . For both parameters we will derive an expression for the asymptotic RE of the ML estimator when comparing equal with unequal cluster sizes. Let $\text{Var}(\hat{\beta}_1|\xi)$ and $\text{Var}(\hat{\sigma}_0^2|\xi)$ denote the variance of the estimator of β_1 and σ_0^2 , respectively, given a design ξ . Let ξ_{equal} denote the design of a study with equal cluster sizes, and let ξ_{unequal} denote a design with unequal cluster sizes. The RE of unequal versus equal cluster sizes for the estimator of the treatment effect, $\text{RE}(\hat{\beta}_1)$, and for the estimator of the intercept variance, $\text{RE}(\hat{\sigma}_0^2)$, are defined as:

$$\text{RE}(\hat{\beta}_1) = \frac{\text{Var}(\hat{\beta}_1|\xi_{\text{equal}})}{\text{Var}(\hat{\beta}_1|\xi_{\text{unequal}})} \quad \text{and} \quad \text{RE}(\hat{\sigma}_0^2) = \frac{\text{Var}(\hat{\sigma}_0^2|\xi_{\text{equal}})}{\text{Var}(\hat{\sigma}_0^2|\xi_{\text{unequal}})} \quad (2)$$

When $\hat{\beta}_1$ in equation (2) is the ML estimator, we can derive an expression for the RE. Let \bar{n} denote the average cluster size of the K clusters in the cluster condition and let w_j be defined as $w_j = (\sigma_0^2 + \sigma_\varepsilon^2/n_j)^{-1}$. For equal cluster sizes, we have $n_j = \bar{n}$ for $j = 1, \dots, K$, and the weight w_j is denoted as w_e . Furthermore, let $\psi = \sigma_\delta^2/\sigma_\varepsilon^2$ be the ratio of the error variance in the control condition and the error variance in the cluster condition. The expression for the RE is as follows (see Appendix A for a proof):

$$\begin{aligned} \text{RE}(\hat{\beta}_1) &= \left(\frac{\sum_{j=1}^K w_j}{K w_e} \right) \times \left(\frac{K w_e \sigma_\delta^2 + n_{K+1}}{\sum_{j=1}^K w_j \sigma_\delta^2 + n_{K+1}} \right) \\ &= \left(\frac{\bar{n} + (1-\rho)/\rho}{\bar{n}} \times \frac{1}{K} \sum_{j=1}^K \left(\frac{n_j}{n_j + (1-\rho)/\rho} \right) \right) \\ &\quad \times \left(\frac{\frac{\psi}{\bar{n}(\rho/(1-\rho)) + 1} + \frac{n_{K+1}}{\bar{n}K}}{\frac{1}{\bar{n}K} \times \sum_{j=1}^K \frac{n_j \psi}{n_j(\rho/(1-\rho)) + 1} + \frac{n_{K+1}}{\bar{n}K}} \right) \end{aligned} \quad (3)$$

The RE for $\hat{\beta}_1$ depends on the ratio of the number of persons in the control condition versus the number of persons in the cluster condition, $n_{K+1}/(\bar{n}K)$, which we will denote as the allocation ratio. It also depends on the ratio of the error variance in the control condition versus the error variance in the cluster condition ψ . The following properties hold for $\text{RE}(\hat{\beta}_1)$, when keeping both the allocation ratio and ψ constant:

1. The RE does not depend on the number of clusters K , but does depend on the intraclass correlation ρ and the distribution of cluster sizes.

2. When we multiply each n_j by a factor $c > 0$, and multiply $(1 - \rho)/\rho$ by the same factor, this does not affect the RE. As a consequence, the minimum of the RE will not depend on \bar{n} , but will only be achieved at a different value of ρ for each possible \bar{n} . Note that the standard deviation of cluster sizes changes by a factor c , whereas the minimum RE remains the same. In examining the effect of cluster size variation on the minimum RE, we will therefore use the coefficient of variation (CV), which divides the standard deviation of cluster sizes by the average cluster size.
3. When $\rho \rightarrow 0$ or $\rho \rightarrow 1$, we have $\text{RE} \rightarrow 1$. For $0 < \rho < 1$, we can show that the $\text{RE} < 1$ (as a result of the Jensen inequality; see [23, p. 72]), implying that equal cluster sizes are optimal.

The RE for $\hat{\beta}_1$ in equation (3) is larger than $\sum_{j=1}^K w_j / (K w_e)$, and decreases as ψ decreases (while keeping ρ constant) and as the allocation ratio $n_{K+1} / (\bar{n} K)$ increases (while keeping K and \bar{n} constant). This implies that, for fixed ρ , as $\psi \rightarrow 0$ and for fixed K and \bar{n} , as $n_{K+1} / (\bar{n} K) \rightarrow \infty$, the RE will approach its minimum, $\sum_{j=1}^K w_j / (K w_e)$. The latter is the RE for the ML estimator of the treatment effect in cluster randomized trials [24]. Thus asymptotically, the loss of efficiency is equal to or smaller than the loss of efficiency for cluster randomized trials. The explanation is that, since there is no clustering in one of the treatment arms, the variation in cluster sizes will only affect the statistical information in one treatment arm. As the allocation ratio increases or ψ decreases, $\text{Var}(\hat{\beta}_1)$ will be dominated more by the sampling variance in the cluster condition and so equation (3) approaches the RE for cluster randomized trials.

The expression for the RE of the ML estimator of σ_0^2 is derived in Appendix B as:

$$\text{RE}(\hat{\sigma}_0^2) = \left(\frac{N \sum_{j=1}^K w_j^2 - (\sum_{j=1}^K w_j)^2}{(N - K) K w_e^2} \right) \times \left(\frac{(N - K) + K \left(1 / \left(\bar{n} \left(\frac{\rho}{1 - \rho} \right) + 1 \right) \right)^2}{(N - K) + \sum_{j=1}^K \left(1 / \left(n_j \left(\frac{\rho}{1 - \rho} \right) + 1 \right) \right)^2} \right) \quad (4)$$

The relative efficiency $\text{RE}(\hat{\sigma}_0^2)$ also satisfies property 1 and 2 of $\text{RE}(\hat{\beta}_1)$. Property 3 does not hold, in that, when $\rho \rightarrow 0$, $\text{RE}(\hat{\sigma}_0^2)$ may become larger than 1 (as will also be seen later on, in the Monte Carlo simulation study). When $\rho \rightarrow 1$, $\text{RE} \rightarrow 1$. That asymptotic REs may become larger than 1 for the variance component estimators, was also found for cluster randomized trials and multicentre trials [25]. Furthermore, when studying the RE for these designs through a Monte Carlo simulation study for small sample sizes, the simulated REs also exceeded 1 [26]. Note that $\text{RE}(\hat{\sigma}_0^2)$ (see equation (4)) does not depend either on the allocation ratio or on the ratio of error variances, ψ . Its expression is the same as that for cluster randomized trials [26].

4. MONTE CARLO INVESTIGATION OF THE RE

Through an extensive Monte Carlo simulation study, we will examine to what extent the asymptotic results on RE hold for numbers of clusters and cluster sizes that are representative of realistic sample sizes. The conditions of the simulation study are similar to those of Candel *et al.* [26]. However, in the context of comparing group with individual treatments, the number of clusters as

Table I. Overview of the conditions of the Monte Carlo simulation.

Factor	Levels
Distribution of cluster sizes*	Unimodal, uniform, bimodal, positively skewed, negatively skewed distribution
Average cluster size [†]	6, 10
Number of clusters	12
Intraclass correlation	$\rho=0.01$ up to 0.30, with steps of 0.01
Ratio of size of control versus cluster condition	0.25, 1, 4
Ratio of error variance in control versus cluster condition	$\psi=0.5, 1, 2$
Estimation method	ML, REML

*More details on the different cluster size distributions are given in Table II.

[†]A larger average cluster size of 10 was considered to be able to examine a larger range of cluster sizes.

well as the cluster sizes are typically smaller (e.g. [4, 6, 8, 12, 19]). This is also true for clinical trials where clustering is induced by the therapist (e.g. [14, 17, 20]). For these smaller samples, it will be useful to examine the accuracy of the asymptotic RE.

4.1. Design of the simulation study

In the simulation study the intercept β_0 was set at 50 and the effect parameter β_1 was set at 5. Since the variance of the effect estimator (see equation (A8)) as well as the variance of the estimator of the intercept variance (see equation (B1)) are asymptotically independent of the fixed regression weights, this could be done without loss of generality. The following factors were systematically varied in the simulation study: (1) the frequency distribution of the cluster sizes, (2) the average cluster size \bar{n} , (3) the size of the intraclass correlation ρ , (4) the ratio of the error variances ψ , (5) the allocation ratio $n_{K+1}/(\bar{n}K)$ and (6) the estimation method. Table I displays the choices made for these factors, the motivation is given in what follows.

4.1.1. Frequency distribution. Five different cluster size distributions were studied: (1) a unimodal, (2) a uniform, (3) a bimodal, (4) a positively skewed and (5) a negatively skewed distribution. Three different cluster sizes, g_a, g_b, g_c , with respective frequencies f_a, f_b and f_c were employed. Details of the cluster size distributions can be found in Table II.

4.1.2. Average cluster size. For the cluster condition we chose an average cluster size $\bar{n}=6$, with a range of 4. For symmetrical distributions this implies that the smallest $n=4$ and the largest $n=8$. Since larger ranges of cluster sizes are not realistic for such \bar{n} , a larger average cluster size of $\bar{n}=10$, was also examined to allow a range of cluster sizes as large as 12. These average cluster sizes are representative of cluster sizes commonly encountered in trials comparing group with individual interventions (e.g. [4–10, 12, 19, 27]), and in trials with therapist-induced clustering in one of the treatments arms (e.g. [13, 15, 17, 20]).

The number of clusters was fixed at $K=12$. This is the representative of the number of clusters that is needed for designs with equal cluster sizes to have sufficient power [18]. As was shown in Section 3, $RE(\hat{\beta}_1)$ (holding ψ and the allocation ratio constant) does not depend on K , and choosing a larger \bar{n} gives the same RE at a smaller value of the intraclass correlation ρ . However, K and \bar{n} were deliberately chosen to be small to check the accuracy of the asymptotic REs.

Table II. Distributions of the cluster sizes in the cluster condition as examined in the Monte Carlo simulation.

Distribution ^{*,†}	Cluster sizes			Range of cluster sizes	Coefficient of variation (CV) of cluster sizes
	g_a	g_b	g_c		
Unimodal	4	10	16	12	0.42
($f_a=3, f_b=6, f_c=3$)	4	6	8	4	0.24
Uniform	4	10	16	12	0.49
($f_a=4, f_b=4, f_c=4$)	4	6	8	4	0.27
Bimodal	4	10	16	12	0.55
($f_a=5, f_b=2, f_c=5$)	4	6	8	4	0.30
Positively skewed	7	10	19	12	0.42
($f_a=6, f_b=4, f_c=2$)	5	6	9	4	0.23
Negatively skewed	4	10	12	8	0.28
($f_a=2, f_b=4, f_c=6$)	3	6	7	4	0.24

* f_a =number of clusters of size g_a (small), f_b =number of clusters of size g_b (medium), f_c =number of clusters of size g_c (large).

† Cluster sizes and cluster frequencies are chosen such that the total number of clusters in the cluster condition is equal to $K=12$, the average cluster size $\bar{n}=6$ or 10, and the ranges of the cluster sizes ($=g_c - g_a$) vary between 12 and 4.

4.1.3. Intraclass correlation. The intercept variance σ_0^2 was set equal to values ranging from 1 to 30, with the error variance in the cluster condition σ_ε^2 simultaneously varying from 99 to 70, to keep the total variance in the cluster condition at 100. As a result, the intraclass correlation in the cluster condition, ρ , ranged by steps of size 0.01 from 0.01 to 0.30, which represents the range of intraclass correlations that are commonly encountered in cross-sectional studies [28, 29].

4.1.4. Ratio of error variances. The expression for $\text{RE}(\hat{\beta}_1)$ depends on the ratio of the error variances, ψ (see equation (3)). Several studies indicate that this ratio varies between 1 and 2 [4, 8, 30]. In the present study these two values will be examined. Since $\text{RE}(\hat{\beta}_1)$ increases as a function of ψ , we will also consider $\psi=0.5$ to obtain a pessimistic estimate of the efficiency loss in this case.

4.1.5. The allocation ratio. The expression for $\text{RE}(\hat{\beta}_1)$ also depends on the allocation ratio, $n_{K+1}/(\bar{n}K)$ (see equation (3)). In several studies the allocation ratio appeared to vary between 1.5 and 0.40 [7–10, 12, 14–16, 19, 27, 30, 31]. As shown by Moerbeek and Wong [18], when assuming equal costs for subjects in the cluster condition and the control condition, and equal cluster sizes, the optimal allocation ratio varies between 1 and 0.33. If costs of subjects in the cluster condition are larger, the optimal allocation ratio is larger. We will therefore consider $n_{K+1}/(\bar{n}K)=4, 1$ and 0.25 to cover the range of values that may be representative of realistic studies.

4.1.6. Estimation method. Initially four different likelihood-based estimation methods were considered: ML estimation and restricted maximum likelihood (REML) estimation, combined with either truncation of negative estimates of the variance components to 0, or without such truncation. Since non-truncation in a substantial number of cases led to non-convergence and also (when convergence was obtained) yielded very extreme values for the parameter estimates, only the truncated estimators were examined further.

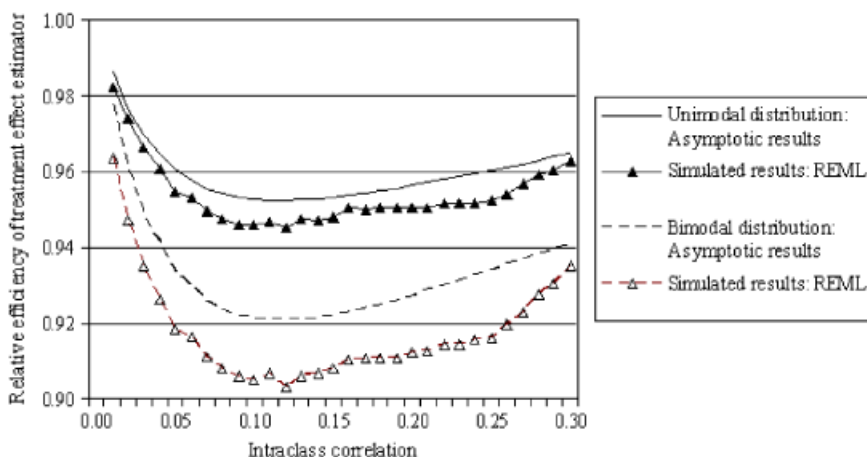


Figure 1. Relative efficiency of the treatment effect estimator for a unimodal and bimodal distribution of cluster sizes, with $\bar{n}=10$, $g_c - g_a = 12$, $\psi=0.5$ and allocation ratio equal to 4. Displayed is also the asymptotic relative efficiency of the ML estimator.

4.2. Simulation procedure

For each of the 2700 simulation conditions 10 000 data sets were generated, each representing the data for 12 clusters consisting, on the average, of 6 or 10 persons. The control condition included either 18, 30, 72, 120, 288 or 480 persons depending on the combination of the average cluster size in the cluster condition and the allocation ratio. The simulations as well as the estimation of the model parameters were performed using version 1.10.0007 of MLwiN [32]. To obtain ML and REML estimates of the parameters, the estimation algorithms ‘Iterative Generalized Least Squares’ and ‘Restricted Iterative Generalized Least Squares’ were employed, respectively. In estimating the model parameters, the convergence criterion was set to 0.001 and the number of iterations was not limited. Based on the obtained estimates for the treatment effect and the random intercept variance, the REs in equation (2) were calculated.

4.3. Results for the treatment effect

Figure 1 shows the simulated RE for $\hat{\beta}_1$ in case $\bar{n}=10$ and, since we are interested in the minimum values for the RE, for $n_{K+1}/(\bar{n}K)=4$ and $\psi=0.5$. Cluster size distributions that give extreme results in terms of RE are shown: the unimodal and the bimodal distribution. Since the ML and REML estimators have comparable REs, only the results for REML are shown. Figure 1 shows that the asymptotic RE and the simulated RE are rather close. In line with the asymptotic RE, the simulated RE never exceeds 1. This implies that unequal cluster sizes are less efficient than equal cluster sizes. For all distributions with $\bar{n}=10$ and range of cluster sizes equal to 12, the RE exceeded 0.92 (the minimum RE for the skewed distributions even exceeded 0.95), except for the bimodal distribution, where the minimum RE approaches 0.90 (see Table III).

When $\bar{n}=6$, the REs are larger, $\text{RE} \geq 0.97$ (see Table III), since the CV of cluster sizes is smaller. For a uniform distribution and $\psi=1$, Figure 2 illustrates that the RE indeed depends upon the allocation ratio as predicted by the asymptotic RE. The more people are assigned to the control condition, the lower the RE becomes. For the same cluster size distribution and an allocation

Table III. Overview of the minimum relative efficiency based on the Monte Carlo simulations, specified for the unimodal, uniform and bimodal distribution of cluster sizes. The allocation ratio $n_{K+1}/(\bar{n}K)=4$ and $\psi=0.5$.*

Features of the cluster size distribution	$K=12, \bar{n}=10, \text{range}=12$			$K=12, \bar{n}=6, \text{range}=4$		
	Bimodal CV=0.55	Uniform CV=0.49	Unimodal CV=0.42	Bimodal CV=0.30	Uniform CV=0.27	Unimodal CV=0.24
Efficiency criterion						
$\text{Var}(\hat{\beta}_1)$	0.90	0.92	0.94	0.97	0.97	0.98
$\text{Var}(\hat{\sigma}_0^2)$	0.84	0.86	0.87	0.92	0.92	0.95

*The allocation ratio $n_{K+1}/(\bar{n}K)=4$ and $\psi=0.5$ were chosen, since these yield the lowest values for the RE in terms of $\text{Var}(\hat{\beta}_1)$. For the RE in terms of $\text{Var}(\hat{\sigma}_0^2)$ this choice was irrelevant.

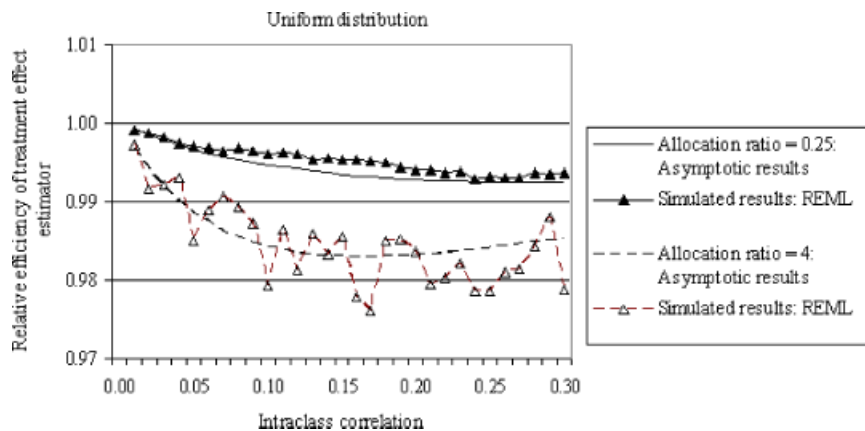


Figure 2. Relative efficiency of the treatment effect estimator for a uniform distribution of cluster sizes, with $\bar{n}=6$, $g_c - g_a=4$, $\psi=1$ and allocation ratio equal to $\frac{1}{4}$ or 4. Displayed is also the asymptotic relative efficiency of the ML estimator.

ratio=1, Figure 3 shows that the higher the error variance in the control condition compared with the error variance in the cluster condition, the higher the RE, also in line with the asymptotic results.

Figures 2 and 3 furthermore demonstrate that also for $\bar{n}=6$ and a uniform distribution of cluster sizes the asymptotic RE and the simulated RE are close. The simulated RE seems to be better approximated by the asymptotic RE, the smaller the CV of the cluster sizes, as illustrated by Figures 1, 2 and 3. The uniform distributions with CV=0.27 show better approximations (Figures 2 and 3) than the unimodal distribution with CV=0.42 (Figure 1). The unimodal distribution in turn shows a better approximation than a bimodal distribution with CV=0.55 (Figure 1). With increasing CV, the asymptotic RE first shifts from somewhat underestimating the simulated RE to overestimating it, and then increasingly overestimates the simulated RE.

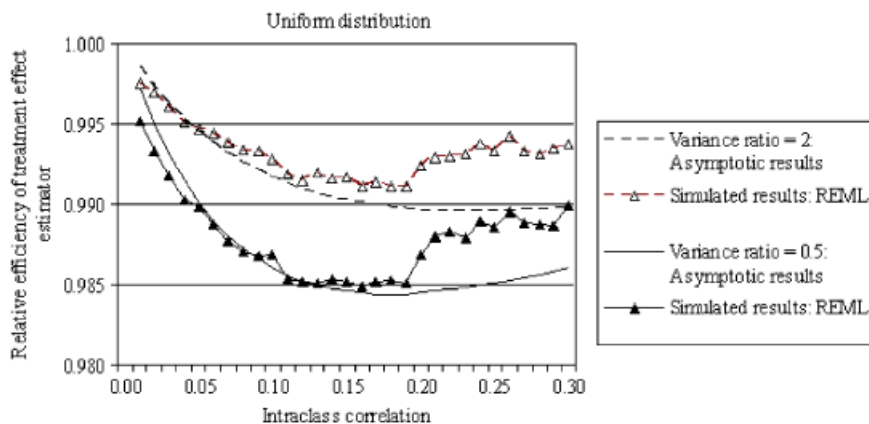


Figure 3. Relative efficiency of the treatment effect estimator for a uniform distribution of cluster sizes, with $\bar{n}=6$, $g_c - g_a=4$, $\psi=0.5$ or 2 and an allocation ratio of 1. Displayed is also the asymptotic relative efficiency of the ML estimator.

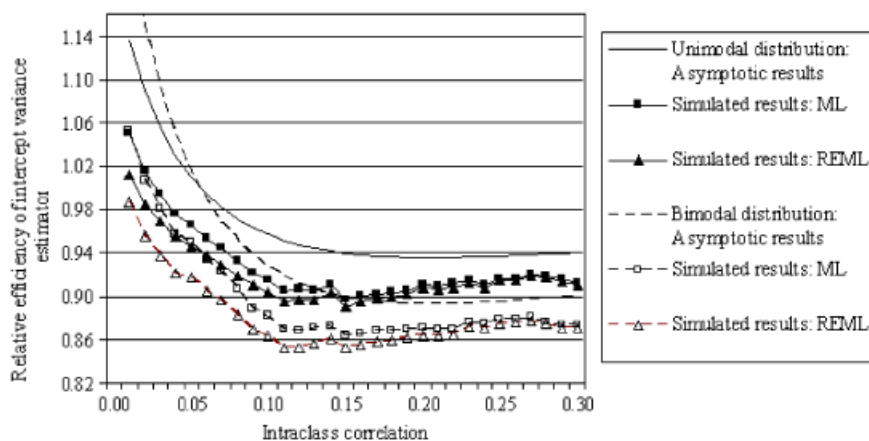


Figure 4. Relative efficiency of the intercept variance estimator for a unimodal and bimodal distribution of cluster sizes, with $\bar{n}=10$, $g_c - g_a=12$, $\psi=0.5$ and allocation ratio equal to 4. Displayed is also the asymptotic relative efficiency of the ML estimator.

4.4. Results for the random intercept variance

As illustrated in Figure 4, the RE for σ_0^2 , in line with the asymptotic results, may become larger than 1 for small intraclass correlations. When estimating the intercept variance, unequal cluster sizes may thus be more efficient. It can also be seen that the simulated RE, particularly in the case of a bimodal distribution and REML, is not very accurately described by the asymptotic RE (for small intraclass correlations, the discrepancy may be larger than 10 per cent). Furthermore, the RE has a lower minimum than the RE for the treatment effect estimator. For the bimodal distribution and $\bar{n}=10$, the RE may become as low as 0.84 (see Figure 4), but for the other distributions it always exceeds 0.86 (see Table III). For skewed distributions, the minimum REs were higher

than 0.91. Again, when $\bar{n}=6$, the REs are larger, $RE \geq 0.92$ (see Table III), since the CV is smaller. As illustrated by Figure 4, for the random intercept variance the ML estimator has a higher RE than the REML estimator.

5. APPLICATION IN PLANNING A TRIAL

When planning a trial in which there is clustering in one of the treatment arms, the results can be used as follows. Let RE denote the relative efficiency of unequal versus equal cluster sizes. Examining equation (A8) in Appendix A shows that there is only one practical way of restoring the efficiency loss when estimating the treatment effect, which is multiplying the number of persons in the control condition as well as the number of clusters in the cluster condition by a factor $1/RE$. It can be shown that it is not the optimal choice but very close to optimal. Optimal here means minimizing the costs involved in a study, for a given precision of estimation.

We will first present a cost function for the design. Let c_c be the costs attached to the persons in the cluster condition, and let c_i be the costs involved with persons in the control condition. Furthermore, let c_g be the extra costs attached to each of the groups in the cluster condition. The total costs of the design, C , can then be defined as:

$$C = n_{K+1}c_i + \bar{n}Kc_c + Kc_g \quad (5)$$

This cost function can also be used if one is interested in minimizing the total number of persons, that is $n_{K+1} + \bar{n}K$, simply by setting $c_c = c_i = 1$ and $c_g = 0$ in equation (5). Since often there is an ideal group size for the cluster condition, we may consider \bar{n} fixed, and we can rewrite equation (5) as:

$$C = n_{K+1}c_i + K(\bar{n}c_c + c_g) = n_{K+1}c_i + Kc_c^* \quad (6)$$

To compensate for the efficiency loss due to unequal cluster sizes in the cluster condition, one would like to increase the number of clusters and persons minimizing this cost function. In Appendix C it is shown that repairing the efficiency loss by multiplying the number of clusters and the number of persons in the control condition by a factor $1/RE$ is not cost-optimal. However, this rule turns out to be almost cost-effective, and therefore is recommended as a simple rule for regaining the efficiency.

When on the other hand, interest focuses on estimating the random intercept variance, only the number of clusters in the cluster condition needs to be multiplied by $1/RE$, since the variance of this estimator only depends on the data of this condition (see equations (B1) and (B2) in Appendix B).

We illustrate the application of these results for a clinical trial by Haugli *et al.* [8], which involves patients with chronic musculoskeletal pain. One group received usual treatment on an individual basis: medication, occasionally supplemented by physiotherapy. The other condition was usual treatment supplemented by a learning program. The learning program consisted of receiving counseling from specially trained health personnel in group sessions. For testing the treatment effect, Moerbeek and Wong [18] provide formulas to calculate the number of groups in the cluster condition, given a fixed and equal group size, as well as the number of persons in the control condition. Taking the estimates from the study of Haugli *et al.* [8] as best guesses as to the true values of the parameters, these formulas can be used to calculate the sample size in case one

would like to replicate this study. Starting from clusters of size $n=8$, a power of 80 per cent, a type I error probability of 5 per cent, $n_{K+1}=102$ subjects are needed for the control condition, and $K=15$ therapy groups are needed for the cluster condition (for details, see [18]).

The present simulation results show that unequal cluster sizes lead to loss of efficiency when estimating the treatment effect β_1 . Since the cluster size distribution is unknown, one may start from a pessimistic scenario assuming a bimodal distribution with, say, $g_a=3$, $g_b=8$ and $g_c=13$, and $f_a=f_c=6$ and $f_b=3$, implying that $CV=0.56$. Assuming furthermore pessimistic choices for $\psi(=0.5)$ and for the allocation ratio ($=4$), the present study suggests that the RE may become as low as 0.90 (see Table III). This loss can be restored in a (near) cost-optimal way by multiplying both the number of groups in the cluster condition and the number of subjects in the other condition $1/RE=\frac{1}{0.90}=1.11$ times, yielding $K=17$ and $n_{K+1}=113$.

Repairing the efficiency loss for the random intercept variance σ_0^2 requires only increasing the number of clusters in the cluster condition. For the same pessimistic scenario, the minimum $RE=0.84$ (see Table III), implying that the number of clusters has to be multiplied by $\frac{1}{0.84}=1.19$. This results in $K=18$ groups of size 8, which still is a modest extension of the original design.

6. CONCLUSIONS AND DISCUSSION

When analyzing data from trials in which treatments have clustering effects, care has to be taken of the dependency between observations within clusters. In case the outcomes within each of the treatment arms are (approximately) normally distributed, linear mixed models provide a useful tool for capturing this clustering. When comparing group therapy with pharmacological treatment or with no treatment at all, or when comparing an individual treatment condition where therapists each treat several patients with a waiting-list condition, clustering may occur in only one of two treatment arms. For these designs Moerbeek and Wong [18] derived formulas for calculating the sample size when testing the treatment effect. In planning such a trial one should, however, also consider the loss of efficiency due to varying cluster sizes. This loss was studied by deriving expressions for the asymptotic RE of unequal versus equal cluster sizes. The extent to which these asymptotic results hold for realistic sample sizes was studied through an extensive Monte Carlo simulation.

The asymptotic RE described the simulated RE for the treatment effect rather adequately. For the random intercept variance, the simulated RE was clearly lower than the asymptotic RE. This was especially true for REML estimation. For the treatment effect the asymptotic results can be used in calculating the efficiency loss, for the intercept variance the simulated relative efficiencies as obtained in this study can be used.

The simulated RE clearly depends on the shape of the cluster size distribution, but also on the variation of cluster sizes. The bimodal distribution and largest CV yield the lowest REs for both efficiency criteria considered. In addition, for the estimator of the treatment effect the simulated RE depends on the ratio of error variances and on the allocation ratio, in a way that is predicted by the asymptotic expression for the RE. There were hardly any differences between ML and REML in the RE of the treatment effect estimator. For the RE of the intercept variance, there was a consistent advantage of the ML estimator over the REML estimator. Then for estimating σ_0^2 , ML appears to be somewhat more robust than REML against varying cluster sizes.

As was illustrated by the simulated REs, the loss of efficiency was modest. This was in line with the asymptotic relative efficiencies being larger than those for cluster randomized trials [24, 26].

The RE of unequal versus equal cluster sizes for the treatment effect estimator exceeds 0.90. It was shown that the almost cost-efficient way of restoring the efficiency level to the level of efficiency of designs with equal cluster sizes was by increasing the number of clusters and the number of persons in the control condition both by $(1/\text{RE}) - 1$, which in practice probably is 11 per cent at most. For the intercept variance estimator, the RE could be as low as 0.84, which can be compensated by only increasing the number of clusters in the cluster condition by $(1/\text{RE}) - 1$, thus by 19 per cent at most. However, note that the exact losses depend on the type of cluster size distribution and on the CV of the cluster sizes. In many cases, the loss of efficiency will be less than the maximum losses found in this study.

Other cluster size distributions involving more than 3 cluster sizes and also involving larger numbers of clusters ($K = 15$ and $K = 16$), as well as models involving other values for the parameters were examined. The results for these examples were in line with the results as obtained in the present Monte Carlo simulation study, thereby supporting the generalizability of the results.

In many intervention studies a binary outcome measure is used. A useful extension of the present study would therefore involve the logistic mixed effects model. It has to be examined whether (approximate) formulas for the asymptotic RE can be derived. These formulas, similar to the present study, could then be tested for their practical utility through a Monte Carlo simulation study.

APPENDIX A: DERIVATION OF THE RE FOR THE ML ESTIMATOR OF THE TREATMENT EFFECT

For $j = 1, \dots, K$, let

$$\mathbf{X}_j = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \cdot & \cdot \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_{K+1} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \cdot & \cdot \\ 1 & 0 \end{bmatrix}$$

Furthermore, for $j = 1, \dots, K + 1$, let the vector of observed scores on the dependent variable be denoted as

$$\mathbf{y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \cdot \\ y_{n_j j} \end{bmatrix}$$

and, let for $j = 1, \dots, K$ the vector of residual scores be given by

$$\boldsymbol{\varepsilon}_j = \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \cdot \\ \varepsilon_{n_j j} \end{bmatrix}$$

and for group $K+1$ we have

$$\boldsymbol{\delta}_{K+1} = \begin{bmatrix} \delta_{1K+1} \\ \delta_{2K+1} \\ \vdots \\ \delta_{n_{K+1}K+1} \end{bmatrix}$$

Let $\boldsymbol{\beta}^T = (\beta_0, \beta_1)$ and $\mathbf{1}$ a vector of n_j ones. In matrix notation, the analysis model (see equation (1)) for the scores of cluster $j = 1, \dots, K$ can be formulated as

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{1} u_{0j} + \boldsymbol{\varepsilon}_j \quad (\text{A1})$$

and for cluster $K+1$ we have:

$$\mathbf{y}_{K+1} = \mathbf{X}_{K+1} \boldsymbol{\beta} + \boldsymbol{\delta}_{K+1} \quad (\text{A2})$$

Let \mathbf{I} denote the identity matrix and \mathbf{J} a matrix with only ones. The variance–covariance matrix of the scores for cluster $j = 1, \dots, K$ can be derived as:

$$\mathbf{V}_j = \sigma_\varepsilon^2 \mathbf{I} + \sigma_0^2 \mathbf{J} \quad (\text{A3})$$

The inverse of this matrix is obtained by applying a result of Searle *et al.* ([33, p. 443]):

$$\mathbf{V}_j^{-1} = \frac{1}{\sigma_\varepsilon^2} \left(\mathbf{I} - \mathbf{J} \left(\frac{\sigma_0^2}{n_j \sigma_0^2 + \sigma_\varepsilon^2} \right) \right) \quad (\text{A4})$$

The inverse variance–covariance matrix for the scores in cluster $K+1$, can be derived as:

$$\mathbf{V}_{K+1}^{-1} = \frac{1}{\sigma_\delta^2} \mathbf{I} \quad (\text{A5})$$

Define $w_j = n_j / (n_j \sigma_0^2 + \sigma_\varepsilon^2)$. The information matrix for the ML estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ is then given by:

$$\text{Inf} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \sum_{j=1}^K \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j + \mathbf{X}_{K+1}^T \mathbf{V}_{K+1}^{-1} \mathbf{X}_{K+1} = \begin{bmatrix} \sum_{j=1}^K w_j + \frac{n_{K+1}}{\sigma_\delta^2} & \sum_{j=1}^K w_j \\ \sum_{j=1}^K w_j & \sum_{j=1}^K w_j \end{bmatrix} \quad (\text{A6})$$

The asymptotic variance–covariance matrix is obtained by taking the inverse, which can be written as:

$$\text{VarCov} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \frac{\sigma_\delta^2}{n_{K+1}} \begin{bmatrix} 1 & -1 \\ -1 & 1 + \frac{n_{K+1}}{\sigma_\delta^2 \sum_{j=1}^K w_j} \end{bmatrix} \quad (\text{A7})$$

From equation (A7), the variance of $\hat{\beta}_1$ can be read as:

$$\text{Var}(\hat{\beta}_1) = \frac{1}{\sum_{j=1}^K w_j} + \frac{\sigma_\delta^2}{n_{K+1}} \quad (\text{A8})$$

Let $w_e = \bar{n}/(\bar{n}\sigma_0^2 + \sigma_\varepsilon^2)$, with \bar{n} the average cluster size in the cluster condition. The relative efficiency (RE) of unequal versus equal cluster sizes, keeping n_{K+1} constant, can now be established as

$$\text{RE}(\hat{\beta}_1) = \left(\frac{\sum_{j=1}^K w_j}{K w_e} \right) \left(\frac{K w_e \sigma_\delta^2 + n_{K+1}}{\sum_{j=1}^K w_j \sigma_\delta^2 + n_{K+1}} \right) \quad (\text{A9})$$

where the left part can be rewritten as

$$\frac{\sum_{j=1}^K w_j}{K w_e} = \frac{\bar{n} + (1 - \rho)/\rho}{\bar{n}} \times \frac{1}{K} \sum_{j=1}^K \left(\frac{n_j}{n_j + (1 - \rho)/\rho} \right) \quad (\text{A10})$$

Employing the Jensen inequality (see e.g. [23, p.72]), it can be seen that the right part between parentheses in equation (A9) is larger than 1. Since the left part between parentheses is the RE for $\hat{\beta}_1$ for cluster randomized trials [24], this implies that $\text{RE}(\hat{\beta}_1)$ in case of trials with clustering in one of the treatment arms is larger.

APPENDIX B: DERIVATION OF THE RE FOR THE ML ESTIMATOR OF THE RANDOM INTERCEPT VARIANCE

The variance components of the cluster and the control condition are estimated for separate parts of the sample, and their estimators are therefore independent. The asymptotic variance–covariance matrix of an ML estimators of the variance components for the cluster condition is given by (see [33, p.89])

$$\text{VarCov} \begin{bmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_\varepsilon^2 \end{bmatrix} = \frac{2}{D} \times \begin{bmatrix} \frac{N-K}{\sigma_\varepsilon^4} + \sum_{j=1}^K \frac{w_j^2}{n_j^2} & - \sum_{j=1}^K \frac{w_j^2}{n_j} \\ - \sum_{j=1}^K \frac{w_j^2}{n_j} & \sum_{j=1}^K w_j^2 \end{bmatrix} \quad (\text{B1})$$

where D is the determinant of the matrix at the right-hand side and can be expressed as [33, p. 90]:

$$D = \frac{N \sum_{j=1}^K w_j^2 - (\sum_{j=1}^K w_j)^2}{\sigma_\varepsilon^4} \quad (\text{B2})$$

The asymptotic RE for the random intercept variance can be derived from equations (B1) and (B2) as being equal to:

$$\text{RE}(\hat{\sigma}_0^2) = \left(\frac{N \sum_{j=1}^K w_j^2 - (\sum_{j=1}^K w_j)^2}{(N-K) K w_e^2} \right) \times \left(\frac{(N-K) + K \left(1 / \left(\bar{n} \left(\frac{\rho}{1-\rho} \right) + 1 \right) \right)^2}{(N-K) + \sum_{j=1}^K \left(1 / \left(n_j \left(\frac{\rho}{1-\rho} \right) + 1 \right) \right)^2} \right) \quad (\text{B3})$$

APPENDIX C: OPTIMAL RESTORATION OF THE EFFICIENCY LOSS DUE TO VARYING CLUSTER SIZES IN CASE OF THE TREATMENT EFFECT ESTIMATOR

Let RE denote the relative efficiency of unequal versus equal cluster sizes. Let f_1 and f_2 denote the multiplication factors of the number of persons in the control condition and the number of clusters in the other condition, respectively. To restore the efficiency loss, equation (A8) shows that these factors should satisfy the following constraint:

$$\frac{\alpha}{f_2} + \frac{\beta}{f_1} = \text{RE} \times (\alpha + \beta) \quad \text{where} \quad \alpha = \frac{1}{\sum_{j=1}^K w_j} \quad \text{and} \quad \beta = \frac{\sigma_\delta^2}{n_{K+1}} \quad (\text{C1})$$

This constraint can be rewritten as:

$$f_2 = \frac{\alpha}{\text{RE} \times (\alpha + \beta) - \beta / f_1} \quad (\text{C2})$$

The cost function that has to be minimized then becomes:

$$f_1 n_{K+1} c_i + f_2 K c_c^* = f_1 n_{K+1} c_i + K c_c^* \alpha / (\text{RE} \times (\alpha + \beta) - \beta / f_1) \quad (\text{C3})$$

Setting the derivative of equation (C3) with respect to f_1 equal to 0, we obtain:

$$n_{K+1} c_i - \frac{K c_c^* \alpha \beta}{(\text{RE} \times (\alpha + \beta) - \beta / f_1)^2} \times \frac{1}{f_1^2} = 0 \Leftrightarrow \frac{n_{K+1} c_i}{K c_c^*} = \frac{\alpha \beta}{(\text{RE} \times (\alpha + \beta) - \beta / f_1)^2} \times \frac{1}{f_1^2} \quad (\text{C4})$$

When the number of persons in the control condition and the number of clusters are chosen under the same cost function as in equation (5), then according to [18, p.2854], we have $(n_{K+1}/K)^2 = w_e \sigma_\delta^2 \times c_c^* / c_i$, and we can rewrite the left-hand side of equation (C4) as follows:

$$\frac{n_{K+1} c_i}{K c_c^*} = \frac{K}{n_{K+1}} \times \left(\frac{n_{K+1}}{K} \right)^2 \times \frac{c_i}{c_c^*} = \frac{K}{n_{K+1}} \times w_e \sigma_\delta^2 = \frac{K w_e}{\sum w_j} \times \sum w_j \times \frac{\sigma_\delta^2}{n_{K+1}} = \frac{K w_e}{\sum w_j} \times \frac{\beta}{\alpha} \quad (\text{C5})$$

Substituting this result in equation (C4), we obtain, after some rewriting:

$$f_1^2 [\text{RE} \times (\alpha + \beta)]^2 - 2 f_1 \text{RE} \times (\alpha + \beta) \beta - \left(\alpha^2 \left(\frac{\sum w_j}{K w_e} \right) - \beta^2 \right) = 0 \quad (\text{C6})$$

Solving for f_1 yields two solutions, one of which minimizes the cost function in equation (C3) (since the second derivative of (C3) turns out to be positive for this f_1):

$$f_1 = \frac{\alpha \times \sqrt{\frac{\sum w_j}{K w_e}} + \beta}{\text{RE} \times (\alpha + \beta)} \quad \text{and thus} \quad f_2 = \frac{\alpha \times \sqrt{\frac{\sum w_j}{K w_e}} + \beta}{\sqrt{\frac{\sum w_j}{K w_e}} \times \text{RE} \times (\alpha + \beta)} \quad (\text{C7})$$

As can be seen, if we have equal cluster sizes, then $\text{RE} = \sqrt{\sum w_j / (K w_e)} = 1$ and thus $f_1 = f_2 = 1$, since no repair is needed. In addition note that $f_1/f_2 = \sqrt{\sum w_j / (K w_e)}$, showing that, since $\sqrt{\sum w_j / (K w_e)} < 1$, relatively more clusters should be added to the cluster condition than persons to the control condition. A problem with the cost-minimizing factors in equation (C7) is that they require knowledge on $\alpha = 1 / \sum_{j=1}^K w_j$ and $\beta = \sigma_\delta^2 / n_{K+1}$, and also on $\sqrt{\sum w_j / (K w_e)}$, which often will not be available. However, although choosing $f_1 = f_2 = 1/\text{RE}$ is not optimal for regaining the efficiency, it can be shown to be highly efficient with costs differing from the costs for the optimal factors by less than 1 per cent in almost all cases.[‡]

REFERENCES

1. Donner A, Klar N. Cluster randomization trials in epidemiology: theory and application. *Journal of Statistical Inference and Planning* 1994; **42**:37–56.
2. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* 1997; **2**:173–185.
3. Baldwin SA, Murray DM, Shadish WR. Empirically supported treatments of type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology* 2005; **73**:924–935.
4. Roberts C, Roberts SA. The design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials* 2005; **2**:152–162.
5. Bauer DJ, Sterba SK, Hallfors DD. Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research* 2008; **43**:210–236.
6. Pals SL, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. *American Journal of Public Health* 2008; **98**:1418–1424.
7. Dannon PN, Gon-Usishkin M, Gelbert A, Lowengrub K, Grunhaus L. Cognitive behavioral group therapy in panic disorder patients: the efficacy of CBGT versus drug treatment. *Annals of Clinical Psychiatry* 2004; **16**:41–46.
8. Haugli L, Steen E, Laerum E, Nygard R, Finset A. Learning to have less pain—is it possible? A one-year follow-up study of the effects of a personal construct group learning programma on patients with chronic musculoskeletal pain. *Patient Education and Counseling* 2001; **45**:111–118.
9. Otto MW, Pollack MH, Gould RA, Worthington III JJ, McArdle ET, Rosenbaum JF. A comparison of the efficacy of clonazepam and cognitive-behavioral group therapy for the treatment of social phobia. *Journal of Anxiety Disorders* 2000; **14**:345–358.
10. Heller-Boersma JG, Schmidt UH, Edmonds DK. A randomized controlled trial of a cognitive-behavioural group intervention versus waiting-list control for women with uterovaginal agenesis (Mayer-Rokitansky-Küster-Hauser syndrome: MRKH). *Human Reproduction* 2007; **22**:2296–2301.
11. Pisinger C, Vestbø J, Borch-Johnsen K, Jørgensen T. Smoking cessation intervention in a large randomised population-based study. The Inter99 study. *Preventive Medicine* 2005; **40**:285–292.

[‡]A proof of this result is, upon request, available from the first author.

12. Telch CF, Agras WS, Rossiter EM, Wilfley D, Kenardy J. Group cognitive-behavioral treatment for the nonpurging bulimic: an initial evaluation. *Journal of Consulting and Clinical Psychology* 1990; **58**:629–635.
13. Roberts C. The implication of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Statistics in Medicine* 1999; **18**:2605–2615.
14. Ladouceur R, Dugas MJ, Freeston MH, Léger E, Gagnon F, Thibodeau N. Efficacy of a cognitive-behavioral treatment for generalized anxiety disorder: evaluation in a controlled trial. *Journal of Consulting and Clinical Psychology* 2000; **68**:957–964.
15. Thompson LW, Gallagher D, Breckenridge JS. Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology* 1987; **55**:385–390.
16. Van Minnen A, Hoogduin KAL, Keijsers GPJ, Hellenbrand I, Hendriks G-J. Treatment of trichotillomania with behavioral therapy or fluoxetine. *Archives of General Psychiatry* 2003; **60**:517–522.
17. Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC. Treatment of atypical depression with cognitive therapy or phenelzine. A double-blind, placebo-controlled trial. *Archives of General Psychiatry* 1999; **56**:431–437.
18. Moerbeek M, Wong WK. Sample size formulae for trials comparing group and individual treatments in a multilevel model. *Statistics in Medicine* 2008; **27**:2850–2864.
19. Calzone KA, Prindiville SA, Jourkiv O, Jenkins J, DeCarvalho M, Wallerstedt DB, Leiwehr DJ, Steinberg SM, Soballe PW, Lipkowitz S, Klein P, Kirsch IR. Randomized comparison of group versus individual genetic education and counseling for familial breast and/or ovarian cancer. *Journal of Clinical Oncology* 2005; **23**:3455–3464.
20. Wampold BE, Serlin RC. The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods* 2000; **5**:425–433.
21. Hox J. *Multilevel Analysis: Techniques and Application*. Lawrence Erlbaum: Mahwah, NJ, 2002.
22. Snijders TAB, Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage: London, 1999.
23. Mood AM, Graybill FA, Boes DC. *Introduction to the Theory of Statistics* (3rd edn). McGraw-Hill: Tokyo, 1974.
24. Van Breukelen GJP, Candel MJJM, Berger MPF. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Statistics in Medicine* 2007; **26**:2589–2603.
25. Van Breukelen GJP, Candel MJJM, Berger MPF. Relative efficiency of unequal cluster sizes for variance component estimation in cluster randomized and multicentre trials. *Statistical Methods in Medical Research* 2008; **17**:439–458.
26. Candel MJJM, Van Breukelen GJP, Kotova L, Berger MPF. Optimality of unequal cluster sizes in multilevel studies with small sample sizes. *Communications in Statistics: Simulation and Computation* 2008; **37**:222–239.
27. Arean PA, Perri MG, Nezu AM, Schein RL, Christopher F, Joseph TX. Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. *Journal of Consulting and Clinical Psychology* 1993; **61**:1003–1010.
28. Smeeth L, Siu-Woon E. Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC trial of the assessment and management of older people in the community. *Controlled Clinical Trials* 2002; **23**:409–421.
29. Parker DR, Evangelou E, Eaton CB. Intraclass correlation coefficients for cluster randomized trials in primary care: the cholesterol education and research trial (CEART). *Contemporary Clinical Trials* 2005; **26**:260–267.
30. Van de Luitgaarden J, Wiers RW, Knibbe RA, Candel MJ. Single-session expectancy challenge with young heavy drinkers on holiday. *Addictive Behaviors* 2007; **32**:2865–2878.
31. Stangier U, Heidenreich T, Peitz M, Lauterbach W, Clark DM. Cognitive therapy for social phobia: individual versus group treatment. *Behaviour Research and Therapy* 2003; **41**:991–1007.
32. Rasbash J, Browne W, Goldstein H, Yang M, Plewis I, Healy M, Woodhouse G, Draper D, Langford I, Lewis T. *A User's Guide to MLwiN* (2nd edn). Centre for Multilevel Modelling, Institute of Education: London, 2000.
33. Searle SR, Casella G, McCulloch CE. *Variance Components*. Wiley: New York, 1992.